

■募集情報

募集番号	2021-1
部署名	先進的音声翻訳研究開発推進センター
インターンシップ課題名	Improving Low-Resource Machine Translation Using Dynamic
インターンシップ内容の概要	<p>Machine translation enables communication between people who speak completely different languages. For language pairs such as French-English and Chinese-English and domains such as news, where data is plentiful, the quality of machine translation is reasonably high.</p> <p>However, language pairs such as Japanese-Hindi or Marathi-English and domains such as educational or medical, the data is scarce and the translation quality is rather low. It is a common practice to use monolingual data and multilingual approaches to improve translation quality in such situations. Our prior research has shown that subword segmentation is an important aspect of low-resource machine translation.</p> <p>Specifically, we proposed a novel segmentation method named Self-Supervised Dynamic Programming Encoding (SSDPE) which uses monolingual data and showed how it leads to more linguistically intuitive subword segmentation which in turn leads to significantly better translation quality compared to other methods.</p> <p>However, this method is rather slow (despite its success) and one of the objectives of this proposal is to make our model more efficient so that it is fast to train and deploy. Furthermore, we plan to work towards a single segmentation model for several languages and then share these models publicly for people to use. Ultimately, we plan to have a single joint model for fast dynamic segmentation and translation and our expectation is that the monolingual training of the segmentation mechanism will benefit machine translation quality. The final goal is to have a single multilingual model for low-resource languages that can give high quality translations as well as linguistically intuitive subword segmentation for better understanding of the working of these models.</p>

■募集情報

課題に関する問い合わせ先	先進的音声翻訳研究開発推進センター 先進的翻訳技術研究室 隅田 英一郎 (連絡先) eiichiro.sumita@nict.go.jp
応募条件	学年：大学院以上 専攻学科等：計算機科学 その他事項：自然言語処理、機械翻訳に関する知見がある方を優先的に採択する。
実施場所	ユニバーサルコミュニケーション研究所（けいはんな）
実施期間	2021/4/1～2022/3/31 の間で 60 日間
受入予定人数	1 人
選考課題	なし
備考	なし